用Excel解決農業研究資料統計分析的方法(五)－變方分析及LSD

農試所農藝組 呂秀英 魏夢麗 呂椿棠

一、前　言

變方分析(ANOVA, analysis of variance)主要用來測驗若干族群平均數之間是否相等。由於我們在從事試驗研究時，往往會施以多處理或多因子來比較各處理平均值間的差異，因此變方分析的使用非常頻繁。當變方分析中的處理變因之F檢定顯著時，表示至少有一對處理的平均值間有顯著差異，然後可再進行多重比較 (multiple comparison) 來確知到底是哪些處理有差異及其差距大小，如使用最小顯著差異性測驗(LSD, least significance difference test)。本篇說明如何利用Excel來進行一些簡單設計的變方分析以及後續的LSD測驗。

二、Excel在變方分析之應用與限制

基於農業上的試驗設計種類很多，變方分析可能會變得很複雜，因此一般而言，欲進行變方分析，最好借助於專用統計分析軟體，如SAS、SPSS等。不過Excel針對較簡單的設計，在「工具」中的「資料分析」裡提供了三種變方分析：選擇「單因子變異數分析」，可進行CRD單因子試驗的變方分析；選擇「雙因子變異數分析：重複試驗」，可進行CRD兩因子試驗的變方分析；選擇「雙因子變異數分析：無重複試驗」，其實是用來進行RCBD單因子試驗的變方分析。但在安裝 Excel 時，通常不會自動安裝此工具。若在「工具」中沒有發現「資料分析」，首先勾選「工具」中的「增益集」，從其「增益功能」內勾選「分析工具箱」即完成安裝。但Excel的工具分析箱並不提供後續的任何多重比較方法，須另外利用輸入公式和函數始能完成。

三、單因子CRD之變方分析的求算步驟及結果解讀

完全隨機設計(CRD, completed random design)是最簡單的設計，適於材料相當純質且環境偏性易控制的實驗室試驗。假設一個含有A、B、C、D四個處理的CRD單因子試驗，原先設計有10個重複，但因資料漏失以致造成不等重複，資料如圖一所列。由於CRD可允許不等重複，此處以不等重複為例說明，至於等重複的情況更為簡單，Excel的操作方法完全一樣。「資料分析」選擇的「單因子變異數分析」，可進行CRD單因子試驗的變方分析，操作步驟如下(圖一)：

1. 指定「輸入範圍」－可用輸入範圍位址或用滑鼠拉選範圍；如本例指定的輸入範圍若包含處理名稱在內(但注意不能將重複標記一起納入)，可直接輸入範圍為相對位址B2:E12。

作者：呂研究員秀英
聯絡電話：04-23302301-125

生物統計

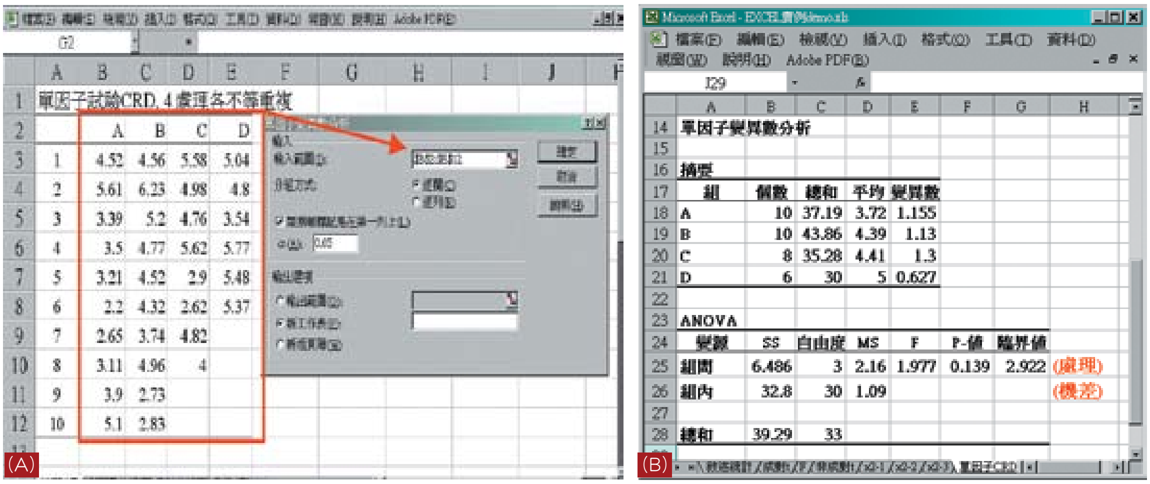
2. 依資料輸入格式選擇「分組方式為逐欄或逐行」－因本例為四個處理分欄輸入，故勾選逐欄。
3. 勾選「類別軸標記是在第一列上」－因本例在指定範圍時，將處理名稱(標記)一起納入，故必須勾選此項，否則毋須勾選。
4. α－內設為0.05，倘毋須變更，則此項可以跳過，否則直接輸入其他數值以改變所設定的顯著水準，如0.01。
5. 在「輸出選項」中勾選分析結果所在位置－放在同一工作表內某處「輸出範圍」內、「新工作表」或「新活頁簿」。
6. 最後按「確定」，即給出輸出結果－Excel會列出摘要和ANOVA結果。在Excel所呈現的變方分析表中的變源即變因(source)，SS則為平方和。變因內的「組內」和「組間」，即「處理」和「機差」(由於Excel屬商業用軟體，商業上對於試驗設計的需求較少，故其在變方分析上的名稱往往與我們所慣用的說法有很大的出入)。可由變方分析中的F值是否大於其臨界值，或直接查看其機率值(P-值)是否小於0.05來決定顯著性，如本例結果顯示(圖一)，F值=1.9773小於臨界值2.9223 (即相當於P-值=0.1386>0.05)，表示該四個處理平均值之間無顯著差異。

四、兩因子CRD之變方分析的求算步驟及結果解讀

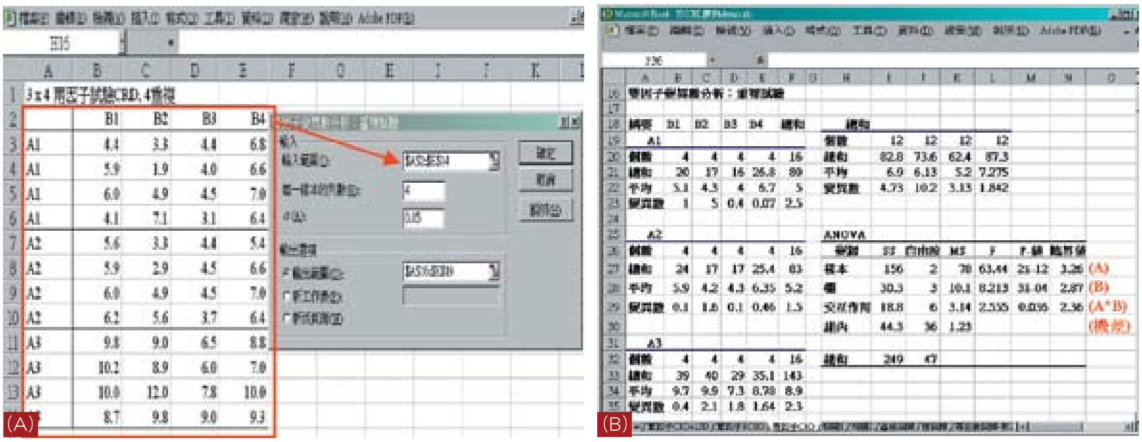
假設一個3×4兩因子CRD試驗(4重複)，其資料輸入到Excel內必須符合一定的格式，如圖二所示，兩因子各為欄和行，而某因子之行內由上而下依次輸入各重複的資料。在「資料分析」內選擇「雙因子變異數分析：重複試驗」，可進行CRD兩因子試驗的變方分析，其操作步驟與前述CRD單因子試驗的變方分析幾乎相同，惟在「輸入範圍」時最好將因子名稱一起指定在內，且於「每一樣本的列數」中輸入重複數，如本例的重複數為4(圖二)，最後按「確定」。在Excel所呈現的變方分析表之變源中，其所謂的「樣本」即被指定於行的因子(如本例的A因子)，而「欄」被指定於欄的因子(如本例的B因子)，「交互作用」即A和B交感效應，「組內」即機差。在此重申，Excel的變方分析輸出結果的中譯標記說明多所不妥，建議最好改為一般慣用稱呼。由圖二結果顯示，本例的A、B主效應皆極顯著存在，且交感效應也顯著存在。

五、單因子RCBD之變方分析的求算步驟及結果解讀

隨機完全區集設計(RCBD, random completed block design)適於田間試驗或有明顯單一偏性因素存在的情況。假設一個含6處理、4區集的RCBD單因子試驗，其資料輸入到Excel內的格式如圖三，由於該試驗含有處理和區集兩種變因，所以可能有兩種輸入方式：處理為欄(行)而區集為行(欄)。在「資料分析」內選擇「雙因子變異數分析：無重複試驗」，可進行RCBD單因子試驗的變方分析，其操作步驟與前述CRD單因子試驗的變方分析幾乎相同，惟若在「輸入範圍」欲將標記一起納入，則務必將區集標記也一起指定後勾選「標記」；否則可不將任何標記納入「輸入範圍」內但不勾選「標記」。必須特別注意，當資料輸入格式是以區集為行而處理為欄，則變方分析表的「行」、「欄」各代表區集效應和處理效應；反之，若資料輸入格式是以



圖一、用Excel進行4處理之不等重複的CRD單因子試驗之變方分析。(A)操作畫面，及(B)輸出結果。



圖二、用Excel進行3×4 CRD兩因子試驗(4重複)之變方分析。(A)操作畫面，及(B)輸出結果。

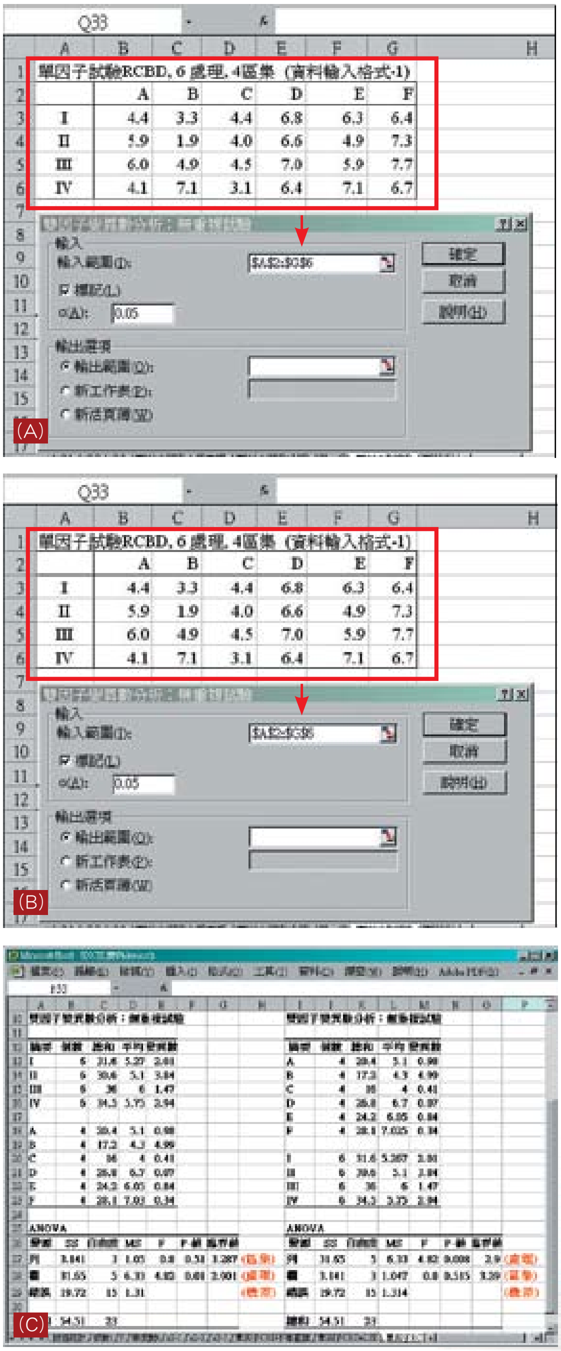
區集為欄而處理為行，則變方分析表的「行」、「欄」各代表處理效應和區集效應。由圖三結果顯示，本例的區集效應不顯著，而處理間存在極顯著差異。

六、LSD測驗之求算步驟及結果解讀

Excel的工具分析箱並不提供任何多重比較方法，但可利用輸入公式和函數的方式來完成。以最廣受應用的LSD為例，首先應求出LSD臨界值，其公式為

$$LSD=t_{\alpha/2,df_{error}}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

式中$t_{\alpha/2,df_{error}}$為α顯著水準(0.05或0.01)下自由度等於變方分析表中之機差自由度的t統計值，MSE為變方分析表中的機差均方，n_i和n_j各為欲比較之兩處理平均值的重複次數。t統計值的計算，可利用Excel的TINV函數，其格式為：TINV(probability, degree_freedom)，



圖三、用Excel進行6處理、4區集之RCBD單因子試驗之變方分析。(A)操作畫面-資料輸入格式1, (B)操作畫面-資料輸入格式2, 及(C)輸出結果-格式1及格式2

probability為雙尾機率值(即若指定顯著水準為0.05，只需直接輸入0.05即可，不必再除以2)，而degree_freedom為機差自由度。而Excel用來開方根的函數為SQRT。當所有處理的重複相等時，兩兩處理間比較的LSD臨界值都是一樣的值，以圖二之RCBD單因子試驗為例，機差自由度的位址在C30，機差均方的位址在D30，在假設0.05顯著水準下，則所有兩兩處理間的LSD臨界值=TINV(0.05,C30)*SQRT(D30*(1/4+1/4))=1.7279。每一對處理平均值間的差值，便可以與其相對應的LSD臨界值做比較，當差值大於LSD臨界值時，表示該兩處理平均值間存在顯著差異。但若重複不等，則兩兩處理間比較的LSD則未必相同，以圖一之CRD單因子試驗為例，0.05顯著水準下各處理之間LSD臨界值的公式分別如下（機差自由度的位址在C27，機差均方的位址在D27）：

A對B: LSD=TINV(0.05,C27)*SQRT(D27*(1/10+1/10))

A對C, B對C: LSD=TINV(0.05, C27) *SQRT(D27*(1/10+1/8))

A對D, B對D: LSD=TINV(0.05, C27) *SQRT(D27*(1/10+1/6))

C對D: LSD=TINV(0.05, C27) *SQRT(D27*(1/8+1/6))